AD_____

GRANT NUMBER DAMD17-96-1-6012

TITLE: Computer-Aided Classification of Malignant and Benign Lesions on Mammograms

PRINCIPAL INVESTIGATOR: Berkman Sahiner, M.D., Ph.D.

CONTRACTING ORGANIZATION: University of Michigan
Ann Arbor, Michigan 48109-1274

REPORT DATE: May 1997

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

19970814 017

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

| 1. AGENCY USE ONLY *(Leave blank)* | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | May 1997 | Annual (1 May 96 - 30 Apr 97) |

**4. TITLE AND SUBTITLE**
Computer-Aided Classification of Malignant and Benign Lesions on Mammograms

**5. FUNDING NUMBERS**
DAMD17-96-1-6012

**6. AUTHOR(S)**
Berkman Sahiner, M.D., Ph.D.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
University of Michigan
Ann Arbor, Michigan 48109-1274

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200*

In the first year of our project, we have made progress in (1) database collection for mammograms containing masses and microcalcifications; (2) segmentation, transformation, and feature extraction from regions of interest on mammograms containing masses; (3) classifier design for the classification of lesions as malignant or benign; and (4) evaluation of algorithms for classification of masses on mammograms.

We have shown that the new image transformation and feature extraction methods improve the mass classification accuracy significantly. We have also shown that, compared to standard feature selection methods, significant improvement can be obtained at the high-sensitivity region of the receiver operating characteristic curve by using a genetic algorithm-based feature selection method. Finally, we have shown that the computerized classification is at least as accurate as the radiologists when tested on a database of 240 mammograms. These results are encouraging steps towards the goal of computer-aided classification of mammographic masses. If these results can be generalized, and computerized results can be shown to improve the classification accuracy of the radiologists in a laboratory setting, it may be possible to incorporate our algorithms in a computer-aided diagnosis package for preclinical testing.

**14. SUBJECT TERMS**
Breast Cancer

**15. NUMBER OF PAGES**
12

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Unlimited |

# FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

N/A Where copyrighted material is quoted, permission has been obtained to use such material.

N/A Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

B.S. Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

B.S. For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

June 1, 1997

PI - Signature Date

# TABLE OF CONTENTS

# INTRODUCTION

Treatment of the breast cancer at an early stage is the most significant means of improving the survival rate of the patients. Mammography is currently the most sensitive method for detecting early breast cancer, and it is also the most practical for screening. However, it is known that a considerable number of lesions visible on the mammograms in retrospect are missed by the radiologists. This can be due to a variety of reasons, including eye fatigue and oversight. Although general rules for the differentiation between malignant and benign lesions exist, in clinical practice, approximately only 15-30% of cases referred for surgical biopsy are actually malignant. We are in the process of developing computer-aided diagnosis (CAD) methods which can provide a consistent and reproducible second opinion to the radiologist for the detection and classification of breast abnormalities.

We are investigating the problem of classifying mammographic lesions as malignant or benign using computer vision, automatic feature extraction, statistical classification, and artificial intelligence techniques. The long-term goal of our research includes the development of an intelligent workstation which would, at the click of a button, provide a consistent and objective second opinion on the probability of lesion malignancy. We hypothesize that such a second opinion would increase the positive predictive value of mammography, reduce the number of unnecessary biopsies without increasing the number of missed carcinomas, and reduce both cost and patient discomfort.

Our efforts are concentrated on the computer-aided classification of two kinds of breast abnormalities, masses and microcalcifications, which are the primary mammographic signs of malignancy. We are investigating computerized extraction of useful features for the differentiation of malignant and benign cases for both abnormalities, and the application of classical statistical classifiers and newly developed paradigms such as neural networks and genetic algorithms for the classification task. Our purposes are to i) improve existing techniques, devise new methods, and identify the preferred approaches for the classification of mammographic lesions, ii) show that computerized classification of mammographic lesions is feasible, and iii) develop a computerized program that can subsequently be shown to improve radiologists' classification of mammographic abnormalities.

# BODY

The progress made so far in the development of the five technical objectives of this project are summarized below. The results obtained using the methods developed in these technical areas are summarized in the discussion of technical objective 5, the evaluation of classification methods. The implications of these results are summarized in the conclusion section.

### Technical Objective 1: Database collection

We have initiated the collection of additional mammograms for our database. The digitizer used to collect the mammograms was changed from the 21 μm CCD based digitizer discussed in the proposal to a 50 μm laser scanning device. The change was made because the CCD digitizer did not pass initial image quality tests, and it had a limited optical density (OD) range. We instead purchased a Lumisys Lumiscan 85 laser film digitizer with an OD range of 0 to 4.0. For the collection of mass cases, the 50 μm pixels were subsampled to 100 μm and archived on optical disks.

The expert mammographer in this project, Dr. Helvie, has been reviewing, categorizing and selecting mammographic cases for digitization based on pathologic finding and tissue density. The regions of interest (ROIs) containing masses and microcalcifications are identified and extracted for the development and evaluation of lesion classification algorithms. We hired a student research assistant in September to digitize and archive the new films and to maintain the mass database. At this point, the student has digitized over 100 films from approximately 20 patients, and we are in the process of incorporating these new films to our evaluation database.

## Technical Objective 2: Feature Extraction for Masses

Segmentation of masses:

We have developed an automated algorithm for segmentation of an ROI into an object region and background tissue We used a pixel-by-pixel clustering algorithm followed by binary object detection for ROI segmentation. We derived several filtered images from the ROI, and used the original and filtered pixel values as the components of the feature vectors in the clustering algorithm. Inclusion of the filtered images made it possible to incorporate neighborhood information into the classification of each pixel. So far, we have applied our clustering algorithm to 255 ROIs containing masses in our database, with satisfactory results. We have not separately attempted to quantify the quality of segmentation. The success of our segmentation algorithm is reflected in the classification results presented under technical objective 5.

Our clustering algorithm, depicted in Fig. 1, is very similar to the K-means algorithm. The goal is to classify pixel $p_i$ as either an object or a background pixel. This is achieved by clustering with feature vector $F_i=[f(1),...,f(L)]$ of length L, where L is the total number of images used in clustering. The algorithm starts by choosing initial cluster center vectors, for the object and the background. Pixels are classified as background or object pixels based on the Euclidean distance between the cluster vector and the cluster center vector. Using this initial classification, two new cluster center vectors are computed. If the new cluster centers are different from the previous ones, the procedure of temporary classification is repeated, otherwise, the clustering is completed.

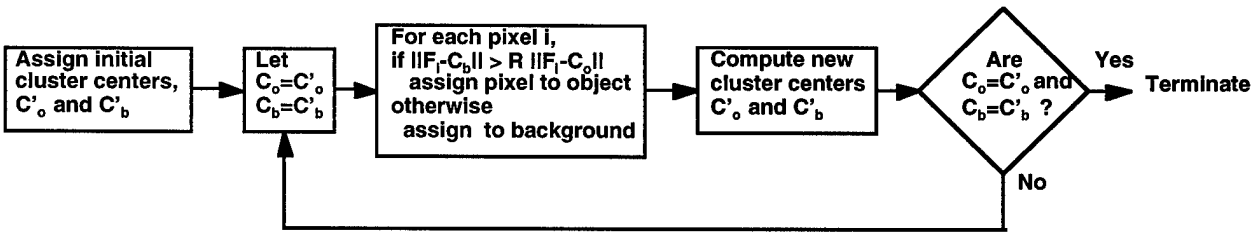


Fig. 1: The block diagram of the clustering algorithm

Fig. 2.a shows an ROI with a spiculated mass. The segmented objects which resulted from the clustering algorithm are shown in Fig. 2.b. After clustering, the largest connected object among all detected objects was selected, filled, and grown in a small region outside its boundary. Fig. 2.c. shows the result of object selection, filling, and object growing applied to Fig. 2.b. Finally, the borders of the grown object were smoothed by using a morphological opening operation. The opening operation for a binary image consists of the successive application of erosion and dilation operations. The final smoothed mass object for the ROI in Fig. 2.a. is shown in Fig. 2.d.

The Rubber Band straightening transform (RBST)

Automatic characterization of the region surrounding a mass is very important in computer aided diagnosis. However, the important features that characterize the mass are directionally dependent and this dependence is affected by the shape of the mass. Commonly used feature extraction methods performed in the Cartesian coordinate system cannot preserve the significant directional information around the mass boundary. As an example, the gradient of the opacity is radially oriented, making it difficult to extract meaningful gradient-based features without preprocessing the image. Similarly, detection of spiculations is complicated by the fact that the search direction for the spiculation changes with the shape of the mass.

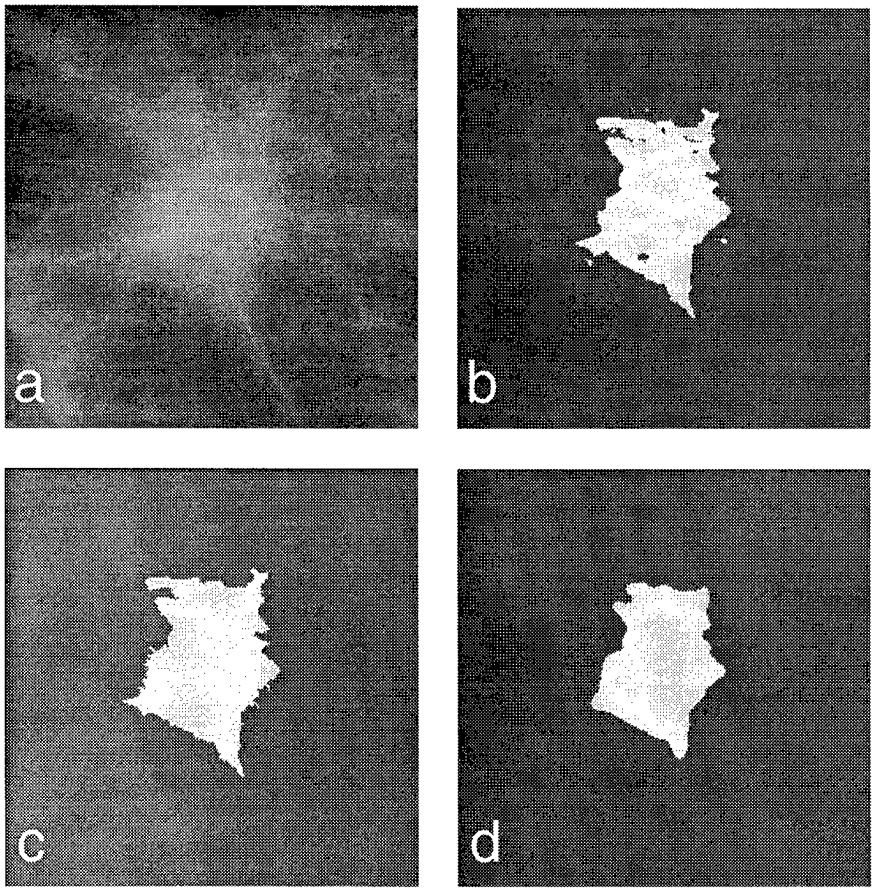


Fig. 2: The result of the mass segmentation algorithm. (a) original ROI, (b) the result of clustering, (c) the result of object selection and growing, (d) the result of morphological filtering

To overcome this problem, we have designed a novel image transformation method, referred to as the rubber-band straightening transform (RBST), to map the band of pixels surrounding the mass onto the Cartesian plane (a rectangular region). In the transformed image, the border of a mass is expected to appear approximately as a horizontal edge, and spiculations are expected to appear approximately as vertical lines. The radially oriented features in the original image will therefore become rectilinear in the transformed image. The RBST facilitates the computerized extraction of important image features. Three main steps in the computation of the RBST, which are edge enumeration, computation of normals, and interpolation are briefly summarized next.

The border pixels of an object form a closed chain, *i.e.,* starting at an arbitrary pixel, it is possible to move along the chain and return to the starting pixel. Conceptually, the edge enumeration algorithm removes pixels, one at a time, from the edge contour of the object, and places the x and y coordinates of each border pixel on an edge enumeration list. Thus, each pixel in the chain is assigned a number, which corresponds to the placement of the pixel in the list. The computation of the normal direction to the object is based on the object shape and the result of the edge enumeration algorithm. For each pixel i in the enumeration list, pixels i+K and i-K, occurring K places before and after pixel i are located in the list, and a normal is drawn to the line joining these two pixels. We have determined that K=12 results in acceptable normal computation. The pixel in row j, column i of the RBST image is defined as the distance-weighted average of the two closest pixels to p(i,j) in the original image, where p(i,j) is the pixel that has distance j along the normal from the border pixel i. The number of columns of the RBST image depends on the number of edge pixels of the segmented mass object, and the number of rows of the RBST image depends on the width of the region desired to be transformed. In this study, we used a 40-pixel-wide region of the ROI surrounding the object to determine the RBST image. An example of an original ROI, segmented mass object, and the RBST image is given in Fig. 3.

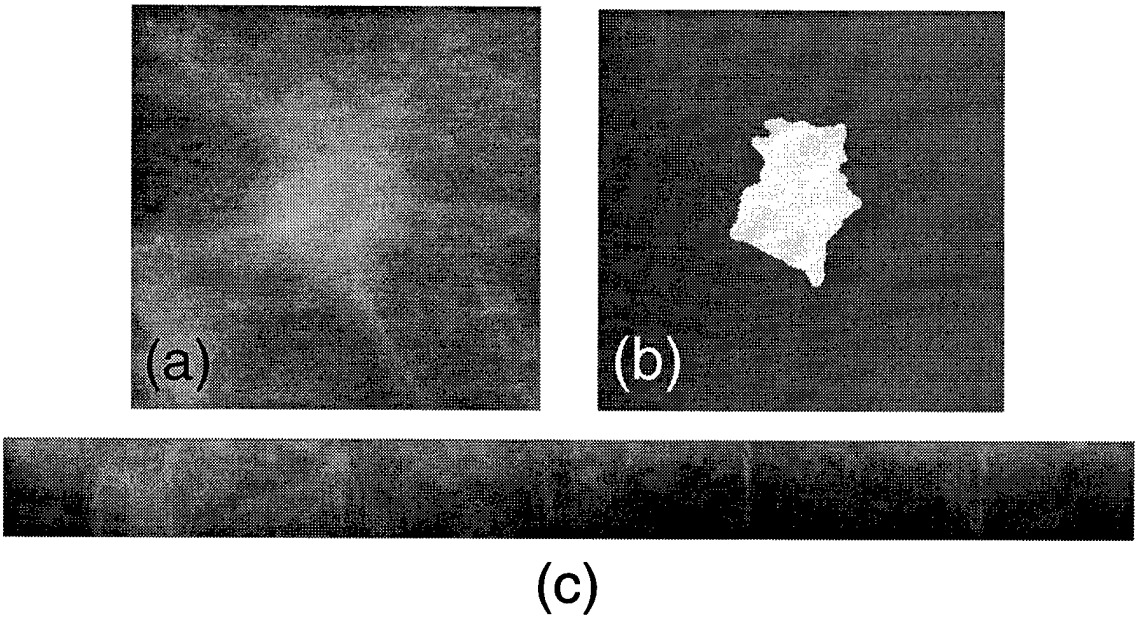


Fig. 3: (a) original ROI, (b) segmented mass object, (c) RBST image

In the past year, we have improved upon the algorithm that computes the RBST, and we have published or submitted articles on its computation and use [1,2,3]. For our database of 255 mammograms, we have shown that features extracted from RBST images yield encouraging classification accuracy. We have also shown that the use of the RBST improves classification accuracy significantly. These results are explained in more detail under technical objective 5.

Texture feature extraction

In the past year, we have investigated the use of texture features extracted from spatial gray-level dependence (SGLD) matrices and run-length statistics (RLS) matrices.

An SGLD matrix can be considered to be a two-dimensional histogram. The (i,j)-th element of an SGLD matrix is the joint probability that gray levels i and j occur in a direction $\theta$ and at a pixel pair distance of d in the image. Texture features, containing information about image characteristics such as homogeneity, contrast, and number and nature of image boundaries were extracted from SGLD matrices.

A gray level run is a set of consecutive, collinear pixels in a given direction which have the same gray level value. A run length is the number of pixels in a run. The RLS matrix describes the run length statistics for each gray level value in the image. RLS texture measures, which summarize the distribution of the RLS matrix elements, were extracted from the RLS matrices.

## Technical Objective 3: Feature extraction for microcalcifications

We did not investigate feature extraction for microcalcifications in the first year of this research project.

## Technical Objective 4: Development of Classifiers

Fisher's linear discriminant with stepwise feature selection

For classification of malignant and benign lesions, we have implemented Fisher's linear discriminant with stepwise feature selection. For a two-class problem, Fisher's linear discriminant projects the multi-dimensional feature space onto the real line in such a way that the ratio of

between-class sum of squares to within-class sum of squares is maximized after the projection. This is the optimal classifier if the features for the two classes have a mutivariate Gaussian distribution with equal covariance matrices.

When the data size is limited, the inclusion of inappropriate features in a classifier may reduce the test accuracy due to overtraining. Therefore, when a large number of features are available, feature selection becomes necessary. Stepwise feature selection in linear discriminant analysis (LDA) is a commonly-used feature selection method. Wilks' lambda, which is defined as the ratio of within-group sum of squares to the total sum of squares, was used as the selection criterion.

Development of a genetic algorithm based high-sensitivity classifier

The cost of missing a malignant lesion is very high in the classification problem. Therefore, it is important to design a classifier with good specificity at high sensitivity. Although stepwise feature selection is a well-established method, it makes no distinction between the goodness of features at high or low sensitivity. In the past year, we have investigated the selection of image features for the design of a high-sensitivity classifier using a genetic algorithm (GA). In designing the high-sensitivity classifier, we used the partial area index under the receiver operating characteristic (ROC) curve $A_{TPF_0}$, which is the average specificity above a sensitivity level $TPF_0$.

A GA comprises a population, which is a set of chromosomes, encoded so that each chromosome corresponds to a possible solution of the optimization problem. The chromosomes consist of genes, which are components of the possible solutions. The chromosomes are allowed to reproduce, exchange genes and mutate. The reproduction probability of each chromosome is related to its ability to solve the optimization problem, i.e., its fitness. By employing $A_{TPF_0}$ as the fitness measure, we were able to selectively identify features which yield good specificity at high sensitivity. We have published an abstract [4] and submitted a journal article [5] on the design of a high-sensitivity classifier.

**Technical Objective 5: Evaluation of classification methods:**

For the development and evaluation of our mass classification methods, we currently use a database of 255 mammograms of patients who had undergone biopsy in the Department of Radiology at the University of Michigan. The database includes 128 biopsy-proven benign masses and 127 biopsy-proven malignant masses. As more mammograms are digitized, they will be entered into the database.

For classification using features extracted from RBST images, texture features were calculated from run-length statistics and spatial gray-level dependence matrices. A total of 320 candidate features at different pixel distances and directions were extracted. Using linear discriminant analysis with stepwise feature selection, 41 features were selected for classification. A leave-one-case-out method was used to train and test Fisher's linear discriminant. The discriminant scores were used as the decision variable for the estimation of the receiver operating characteristic (ROC) curve.

Figure 4 shows the distribution of the test discriminant scores, and Figure 5 shows the resulting ROC curve. The area $A_Z$ under the ROC curve was 0.92. If the decision threshold is chosen properly, more than 30% of benign masses could be correctly classified with no missed malignant masses. These preliminary results are very encouraging.

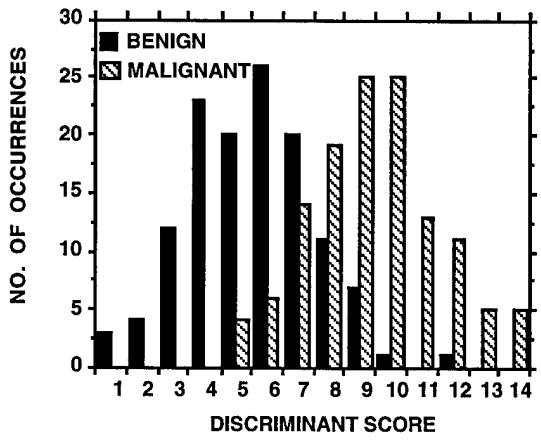


Figure 4: The distribution of the computer test scores

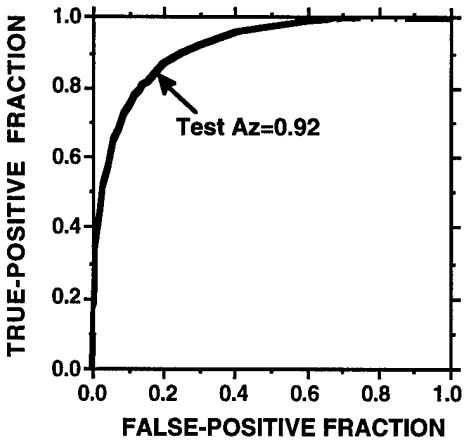


Figure 5: The ROC curve.

To compare the effectiveness of the features extracted from the RBST images with the effectiveness of those extracted from the region surrounding the mass, or from an ROI containing the mass, the same set of 320 texture features were extracted separately from two regions, which were: R1, a 256X256 ROI containing the mass, and R2, a 40-pixel-wide region of the ROI surrounding the segmented mass. The RBST representation was called R3. Figure 6 shows the test ROC curves obtained by the leave-one-case-out method. The area $A_Z$ under the ROC curve was 0.83, 0.85, and 0.92 for R1, R2, and R3 respectively. The difference between classification results using R1 and R3, as well as R2 and R3 were statistically significant ($p<0.05$). These results show that texture features extracted from the RBST images are significantly more effective.

We have used the same set of 320 texture features for feature selection in a GA for the design of a high-sensitivity classifier. The $TPF_0$ used for defining the fitness function was 0.95. Figure 7 compares the resulting ROC curves for Fisher's classifier with stepwise feature selection and the GA-based high-sensitivity classifier. It is observed that the GA-based classifier is superior to Fisher's classifier at high-sensitivity, although the overall area under the ROC curve is lower. The difference between the two classifiers were significant for $TPF>0.90$ ($p<0.05$). This result shows that it is possible to improve upon the stepwise feature selection for the design of a high-sensitivity classifier.

We have also performed observer studies using a 240-mammogram subset of our database [6]. The purpose of the observer study was to compare the classification accuracy of the observers to that of the computer algorithm. Six board-certified, ACR-accredited radiologists were asked to rate their confidence that the mammogram contained a malignant mass on a scale of 1 to 10. The case order was randomized for each observer and the reading time was unlimited. The confidence ratings were analyzed by ROC methodology. The average Az value of the radiologists (obtained by averaging "a" and "b" parameters in ROC analysis) was 0.86. This result also highlights that the classification accuracy attained by our computerized method is at least comparable to that of the radiologists.

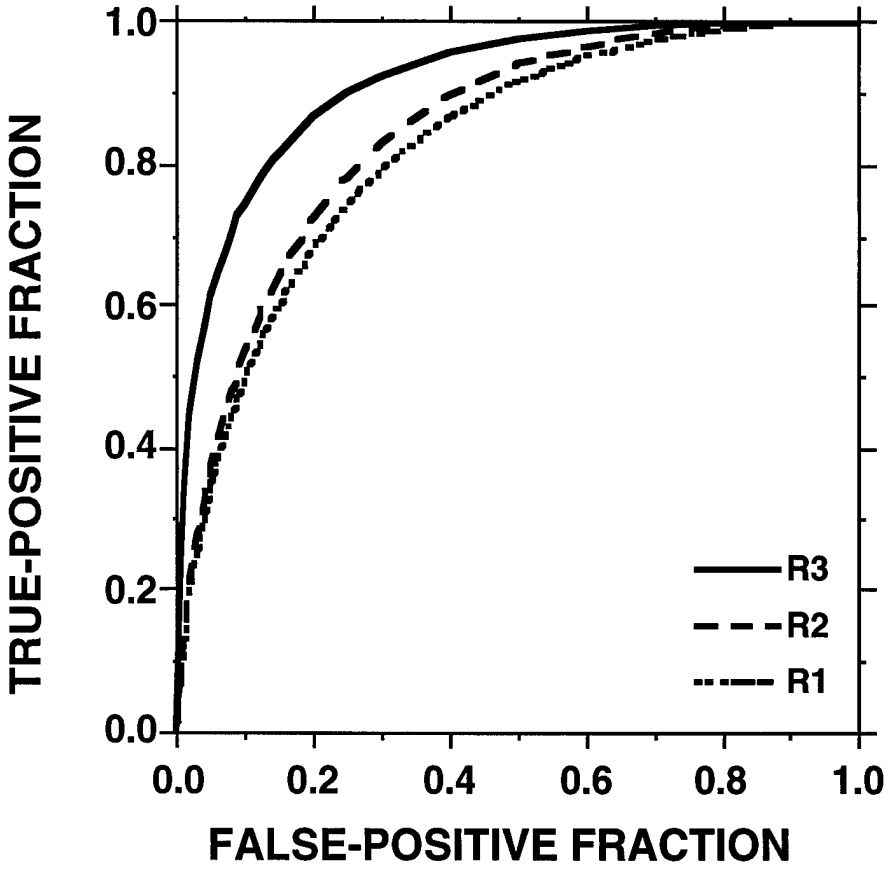


Fig. 6: ROC curves for three image representations using texture features

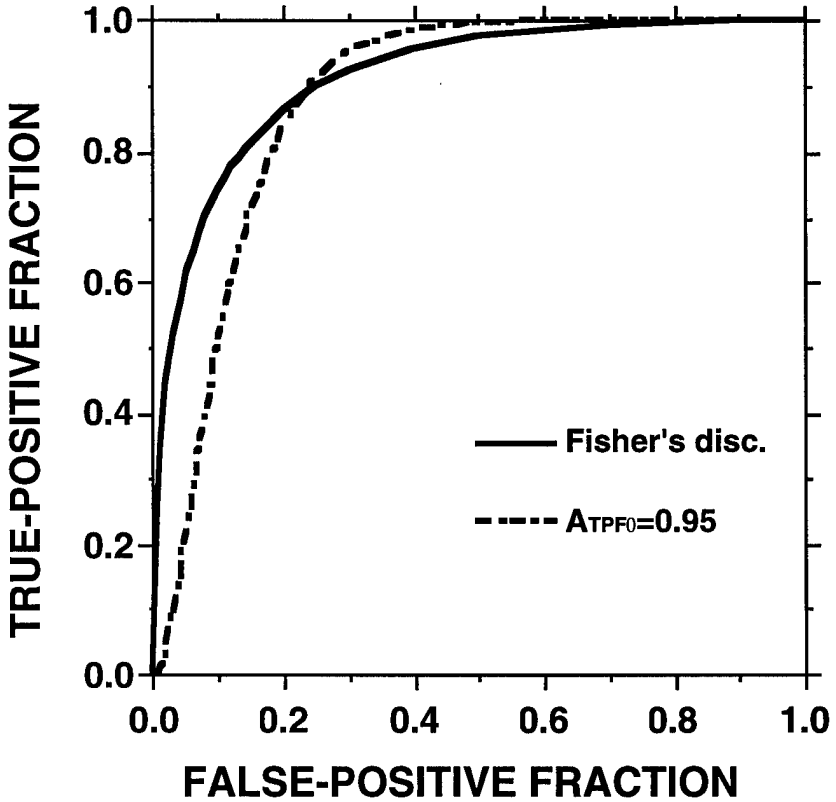


Fig. 7: ROC curves for Fisher's linear discriminant with stepwise feature selection and GA-based high-sensitivity classifier

## CONCLUSION

In the first year of the project, we have made significant progress in four of the five major technical objectives in our proposal. Despite the fact that we have had to change our digitizer after the beginning of the project, we were able to digitize over 100 films with our new digitizer. We have improved upon our existing methods and designed new techniques for the segmentation, transformation, and feature extraction from regions of interest containing masses on mammograms. We have investigated existing classification techniques and designed a new high-sensitivity classifier for the classification of lesions as malignant and benign. We have evaluated our mass characterization algorithms using a database of 255 mammograms.

The results of the evaluations are encouraging. We have shown that the RBST improves the mass classification accuracy significantly. We will therefore pursue the RBST further, and investigate additional features that can be extracted from RBST images. We have also shown that, compared to standard feature selection methods, significant improvement can be obtained at the high-sensitivity region of the ROC curve by using a GA-based feature selection method. Finally, we have shown that the computerized classification at least as accurate as the radiologists when tested on a database of 240 mammograms. The improvement in the classification accuracy when the radiologists are aided by the computer classification scores will be evaluated next year.

## REFERENCES

[1] B. Sahiner, H.-P. Chan, N. Petrick, M.A. Helvie, C.P. Paramagul, and J.S. Newman, "Classification of Mammographic Masses Using Texture and Gradient Features from Transformed Images," *Radiology,* 1996 , 201(P):256 (Abstract).

[2] B. Sahiner, H.-P. Chan, N. Petrick, M.M. Goodsitt, and M.A. Helvie, "Characterization of masses on mammograms: Significance of the use of the rubber band straightening transform," *Proc. SPIE Medical Imaging (in Press),* February 1997.

[3] B. Sahiner, H.-P. Chan, N. Petrick, Mark A. Helvie, and M.M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis," *Medical Physics, (submitted),* 1997.

[4] B. Sahiner, H.-P. Chan, N. Petrick, M.A. Helvie, D.D. Adler, and M.M. Goodsitt "Classification of Malignant and Benign Breast Masses: Development of a High-Sensitivity Classifier Using a Genetic Algorithm," *Radiology,* 1996 , 201(P):257 (Abstract).

[5] B. Sahiner, H.-P. Chan, N. Petrick, Mark A. Helvie, and M.M. Goodsitt, "Computerized classification of malignant and benign masses on mammograms: A high-sensitivity classifier based on a genetic algorithm," *Medical Physics, (submitted),* 1997.

[6] H.-P. Chan, B. Sahiner, N. Petrick, T.E. Wilson, M.A. Roubidoux, and D.D. Adler, "Observer Performance Study of Radiologists' Reading of Mammographic Masses and Comparison with Computerized Classification," *Radiology,* 1996 , 201(P):370 (Abstract).